CHIL data for the NIST Rich Transcriptions 2006 evaluations

Djamel MOSTEFA

Evaluations and Language resources Distribution Agency (ELDA)

http://www.elda.org

Plan

- CHIL Data collection
- Evaluated technologies
- CHIL data for RT'06
- Labelling
- Evaluation packages
- Conclusion

Data collecting sites (1/2)

- Non interactive seminars
  - UKA, Germany
  - ITC-irst, Italy

ITC

UKA

Data collecting sites (2/2)

Interactive seminars

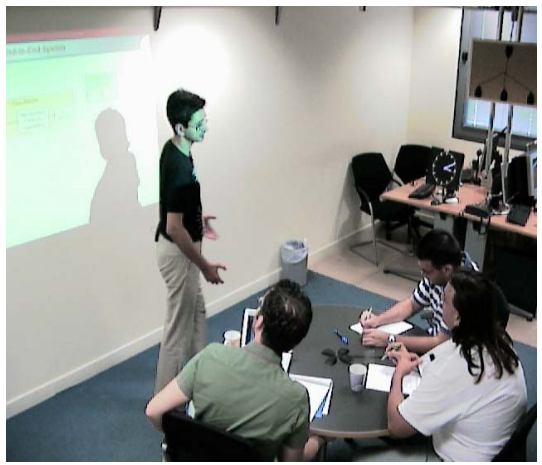
- AIT, Greece
- UPC, Spain

scripted

- IBM, USA

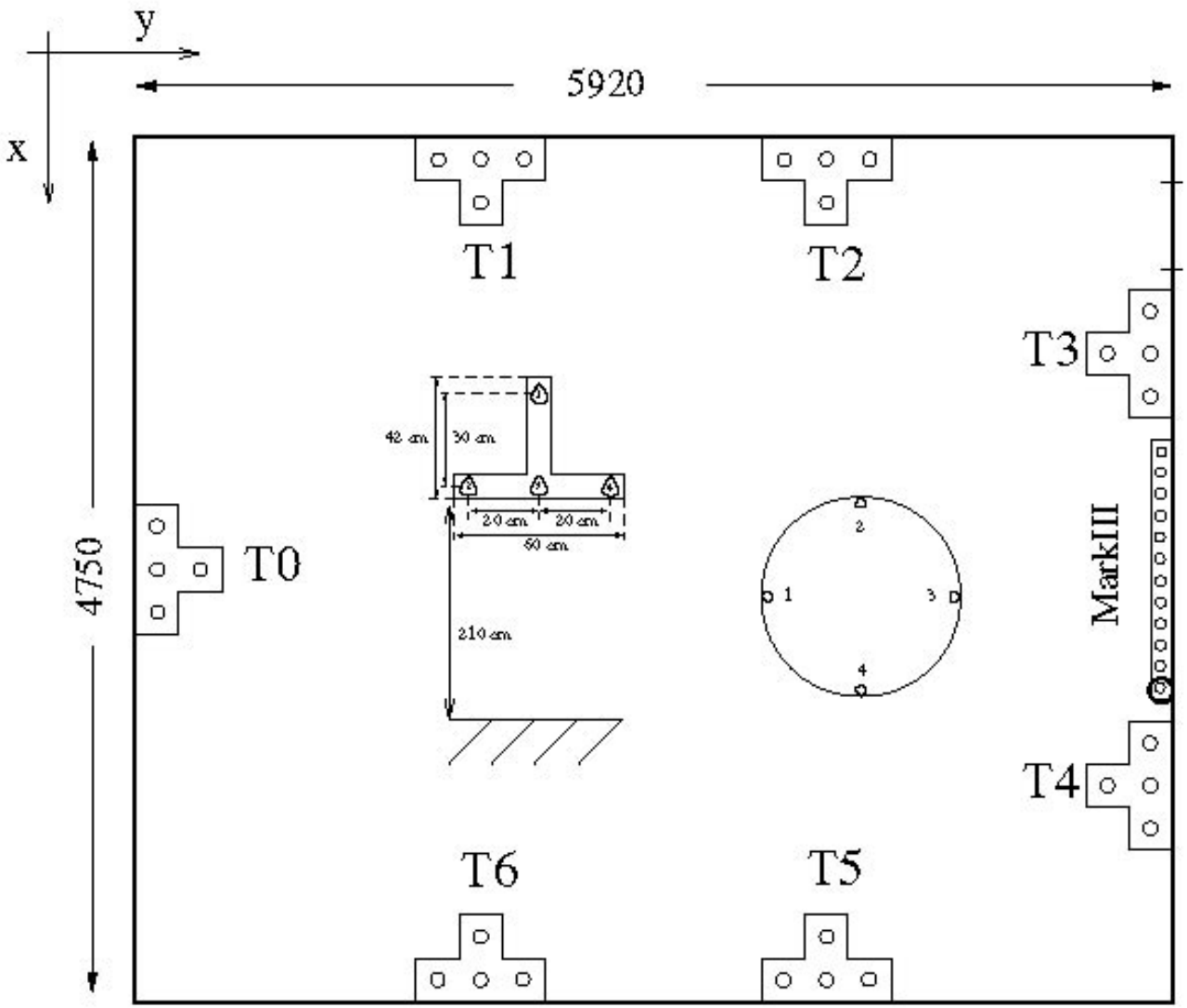
UPC

IBM

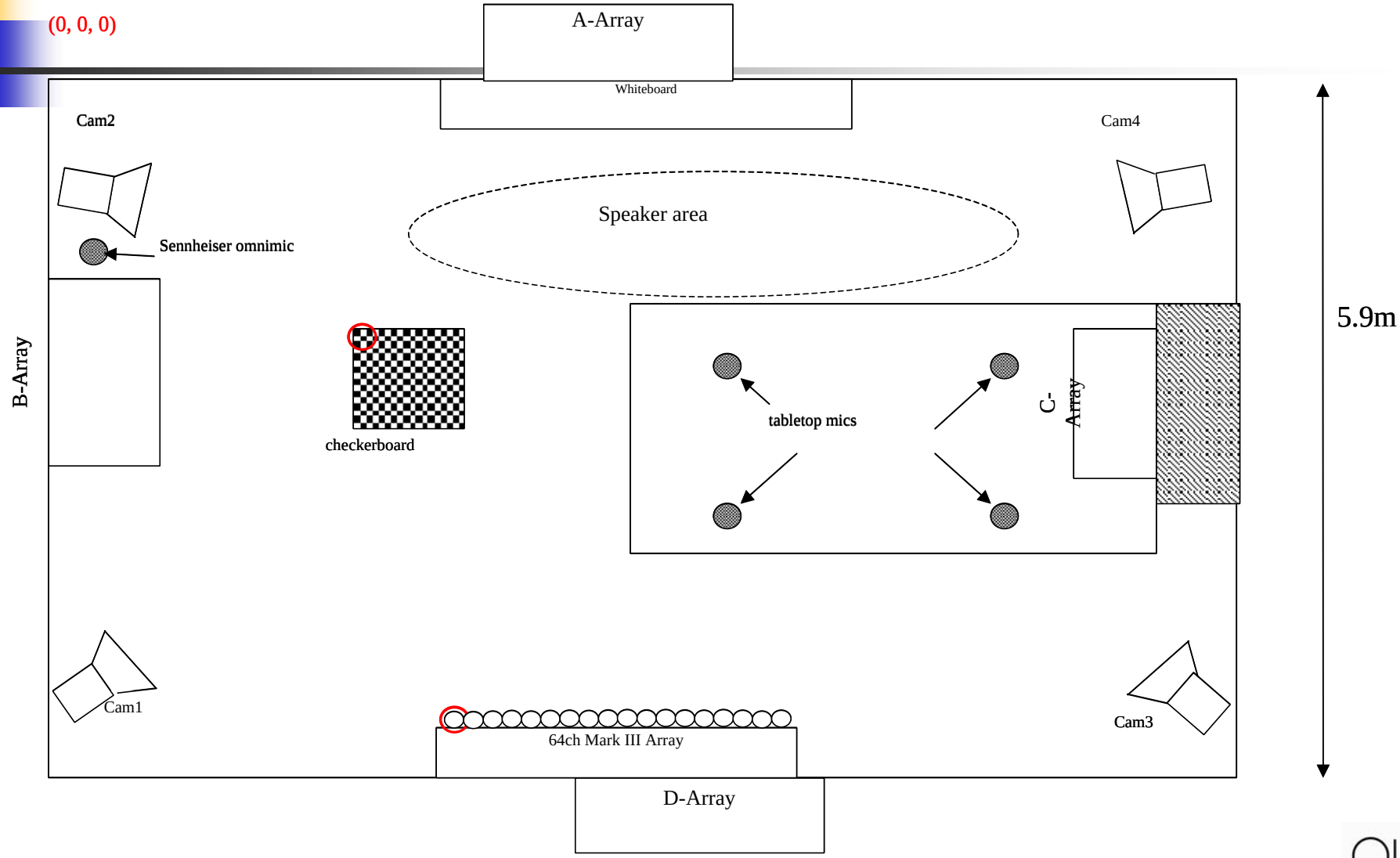


AIT

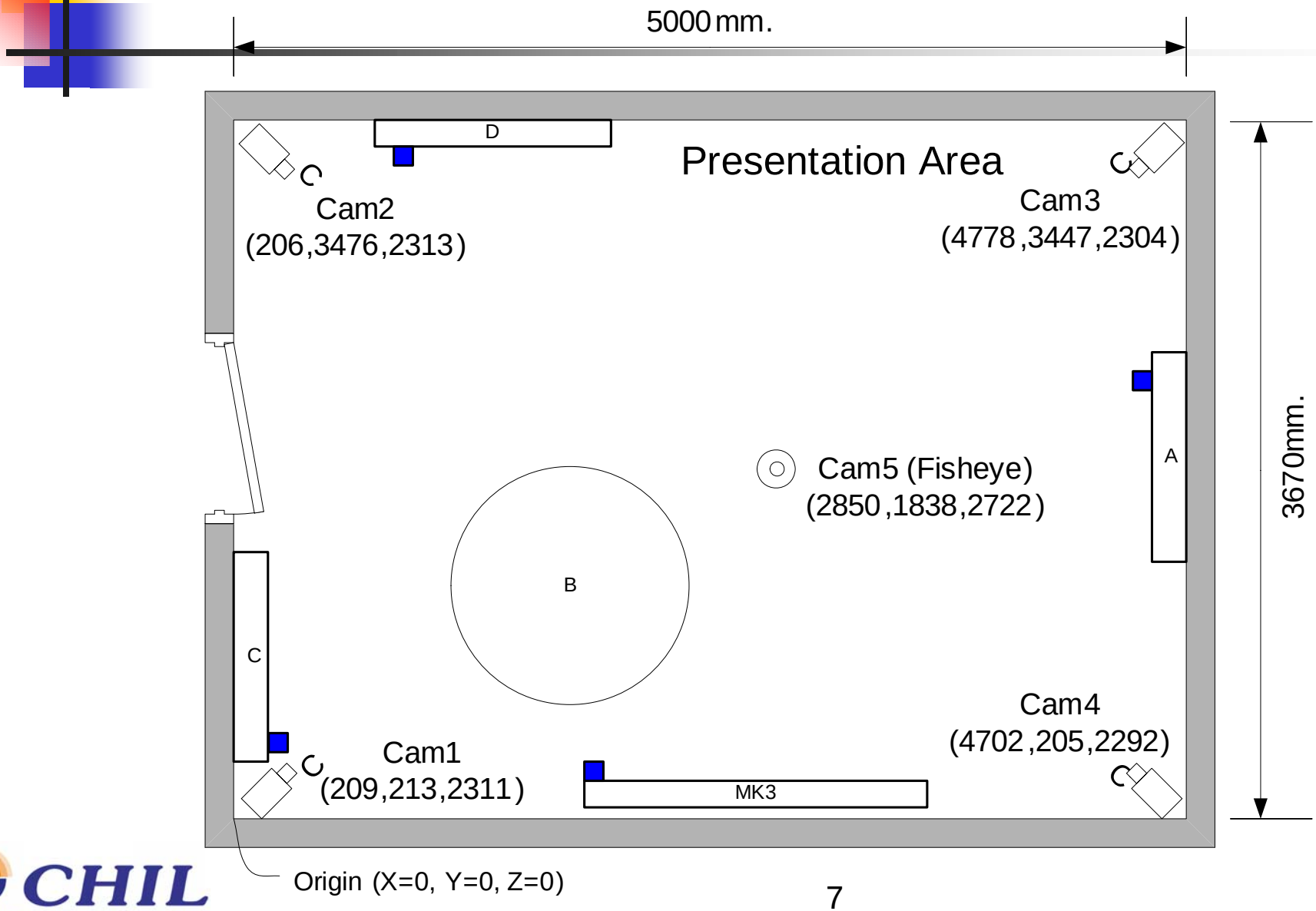
ITC room


y
x
5920
4750
T0
T1
T2
T3
T4
T5
T6
MarkIII
42 cm
30 cm
20 cm
20 cm
60 cm
210 cm
1
2
3
4

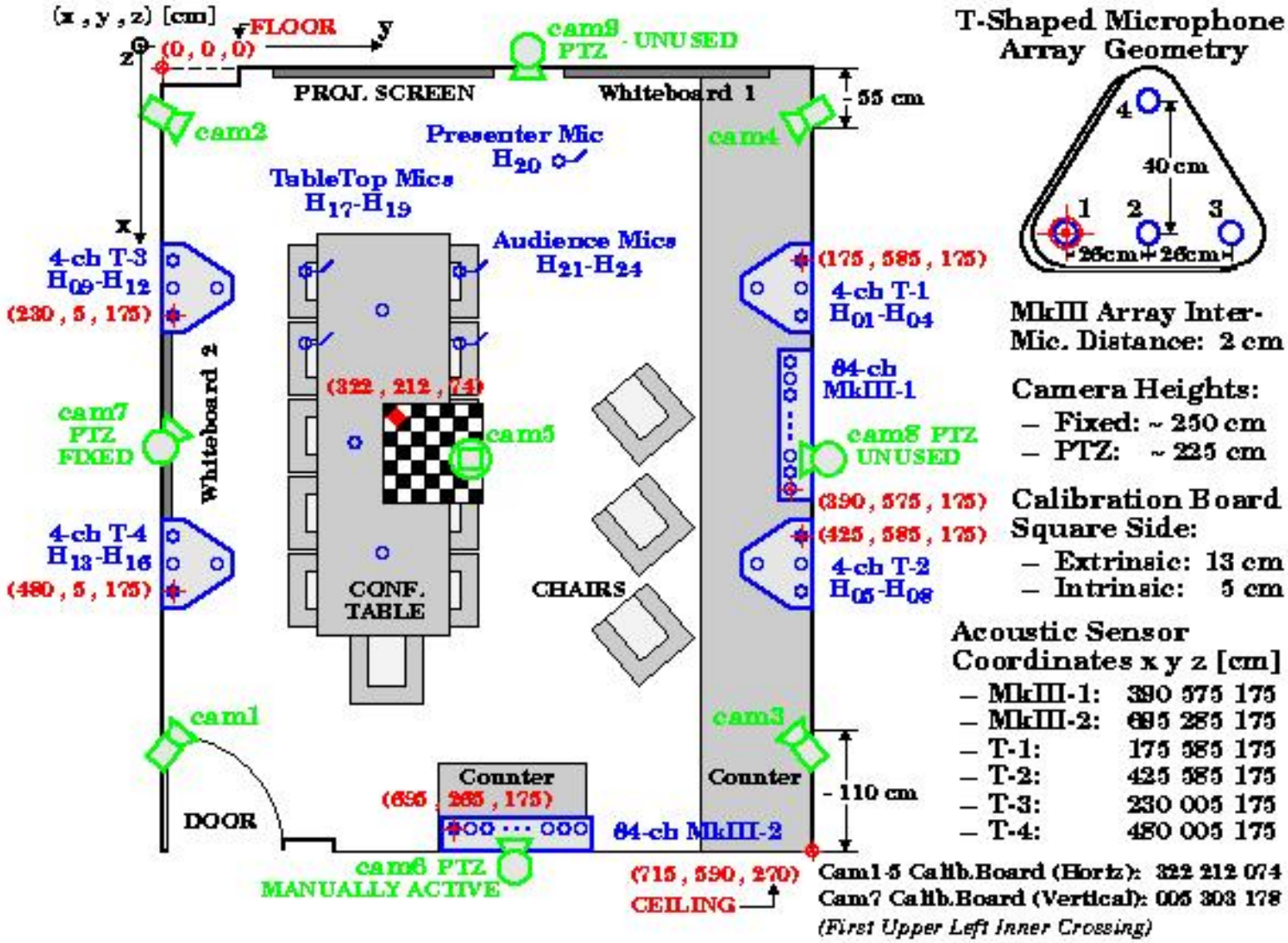
UKA room


(0, 0, 0)
A-Array
Whiteboard
Cam2
Cam4
Speaker area
Sennheiser omnimic
B-Array
checkerboard
tabletop mics
C-Array
5.9m
Cam1
64ch Mark III Array
Cam3
D-Array
7.1m

AIT room setup



IBM room



T-Shaped Microphone Array Geometry

MkIII Array Inter-Mic. Distance: 2 cm

Camera Heights:
- Fixed: ~ 250 cm
- PTZ: ~ 225 cm

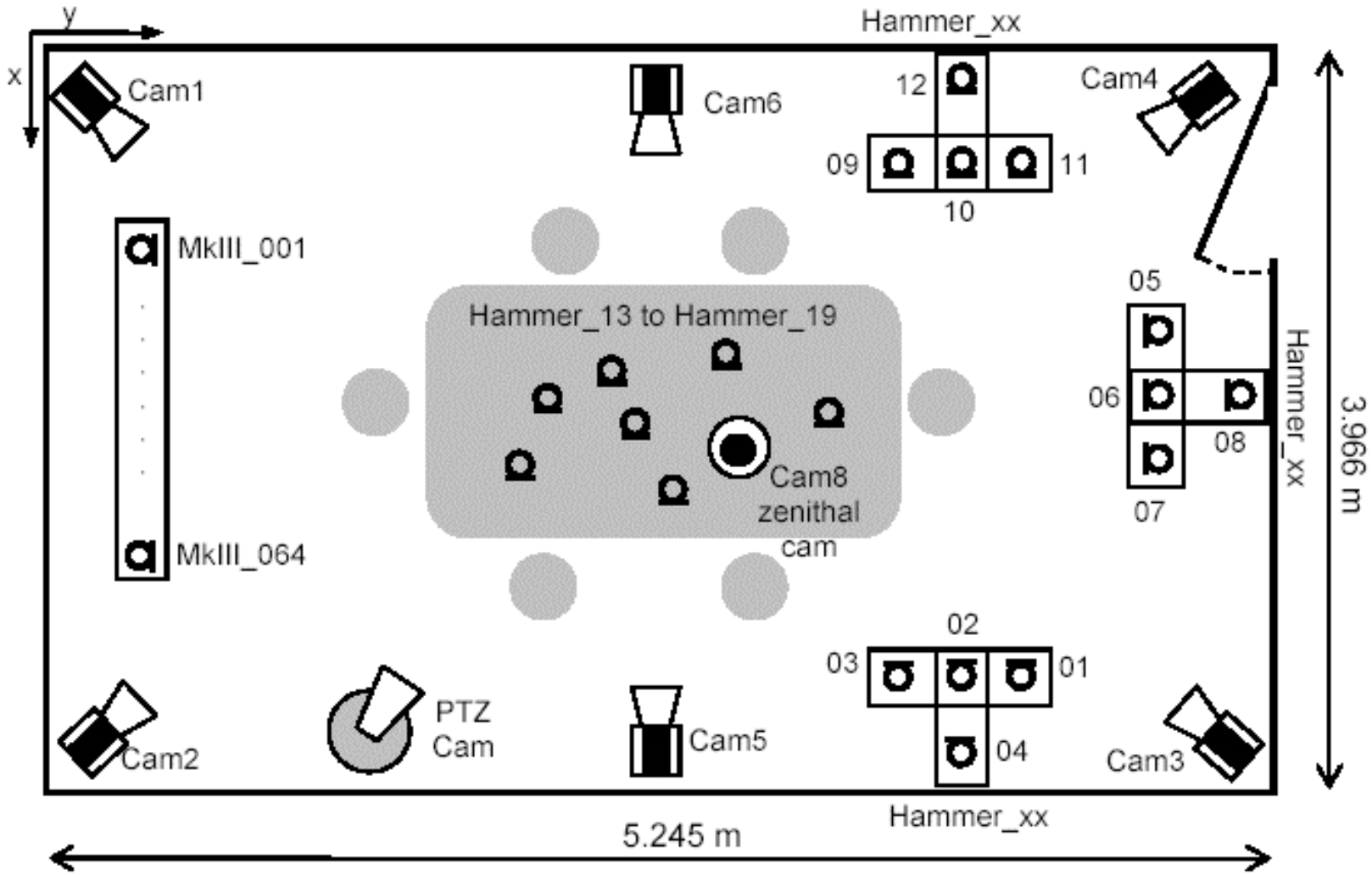
Calibration Board Square Side:
- Extrinsic: 13 cm
- Intrinsic: 5 cm

Acoustic Sensor Coordinates x y z [cm]

– MkIII-1:	390 575 175
– MkIII-2:	695 285 175
– T-1:	175 585 175
– T-2:	425 585 175
– T-3:	230 005 175
– T-4:	480 005 175

Cam1-5 Calib.Board (Horiz): 322 212 074
Cam7 Calib.Board (Vertical): 005 303 178
(First Upper Left Inner Crossing)

UPC room


y
x
Hammer_xx
Cam1
Cam6
12
Cam4
09
11
10
MkIII_001
Hammer_13 to Hammer_19
05
06
08
07
Hammer_xx
3.966 m
Cam8
zenithal
cam
MkIII_064
02
03
01
04
PTZ
Cam
Cam5
Cam2
Cam3
Hammer_xx
5.245 m

Content of a seminar

- Video data
  - 4 or 5 video sequences (JPEG files at 15 to 30 fps)
- Audio data
  - hammerfall channels
    - CTMs, tabletops, T-shape arrays, lapel
    - 44kHz, 24 / 16 bits resolution
  - markIII array
    - 64 channels, 44kHz, 24-bits resolution

Evaluated technologies

- **Vision technologies**
  - Face & Head tracking CLEAR'06
  - Visual Person Tracking CLEAR'06
  - Visual Person Identification CLEAR'06
  - Head Pose Estimation CLEAR'06
- **Audio technologies**
  - Automatic Speech Recognition RT'06
  - Acoustic Person Tracking (in space) CLEAR'06
  - Acoustic Speaker Identification CLEAR'06
  - Speech Activity Detection / speaker diarization RT'06
  - Acoustic Event Detection and Classification CLEAR'06
- **Multimodal technologies**
  - Multimodal Person Identification CLEAR'06
  - Multimodal Person Tracking CLEAR'06
- **Content processing**
  - Questions Answering CHIL'06
  - Automatic Summarization CHIL'06

CHIL Training data

- Training data available for UKA seminars only
  - 2003, 2004 seminars
- No training data for the other sites
  - All data collected in 2005 is used for the evaluations

Devset and evalset

	CHIL 06 Devset	CHIL 06 Eval set
UKA	17 * 5 min	12 * (5 + 5 min)
ITC	1 * 15 min	2 * 5 min
AIT	1 * 15 min	4 * 5 min
IBM	1 * 15 min	4 * 5 min
UPC	1 * 15 min	4 * 5 min
TOTAL	**145 min**	**190 min**

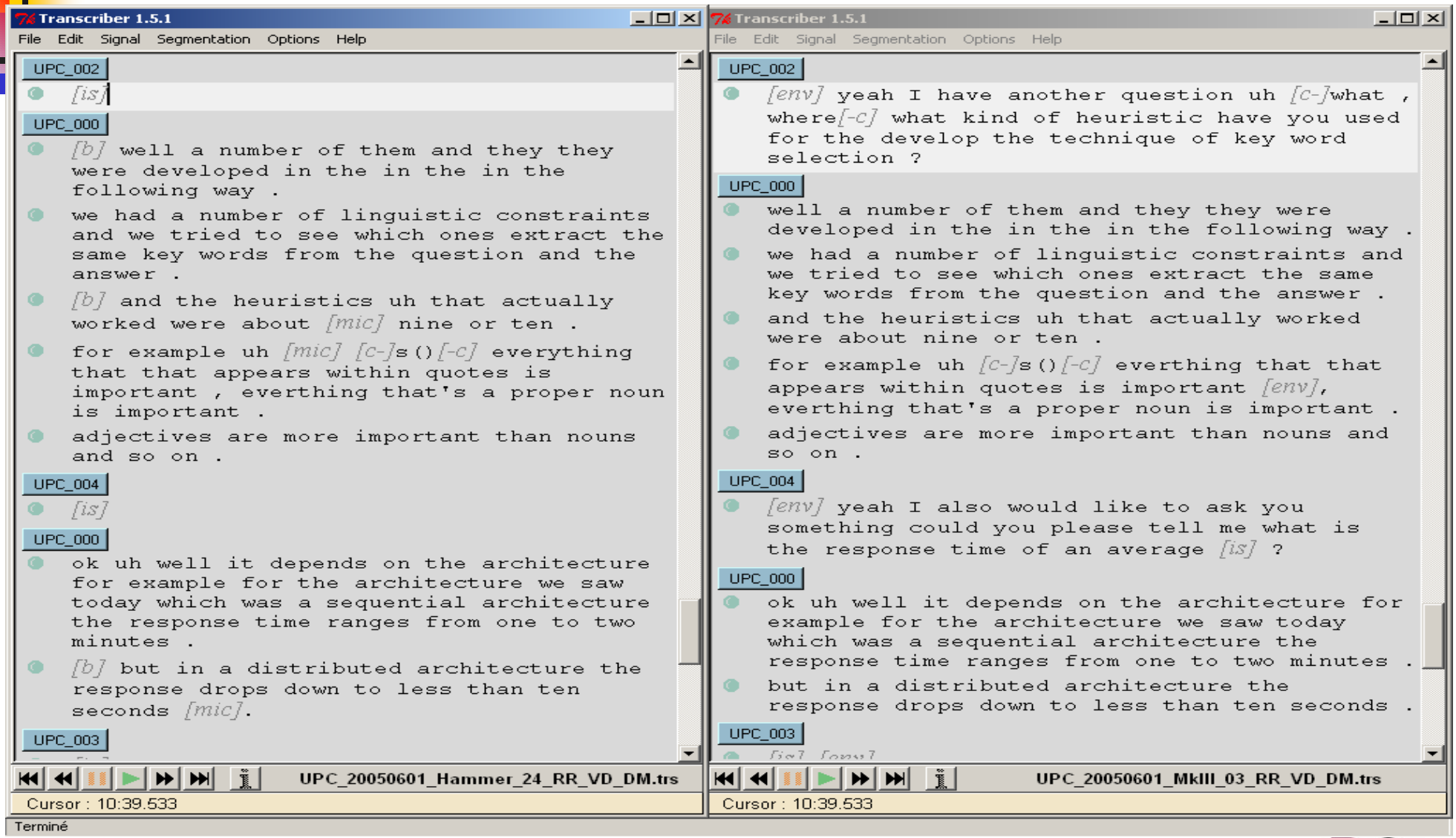
Eval data statistics

- 38 x 5 min from 26 seminars
- 25k running words
- 2.6k lexicon
- 74 speakers
  - 7 f / 67 m
  - Non native English speakers

Transcriptions

- 2 transcriptions are produced:
  - Close-Talking Microphone of the presenter
  - One channel from the microphone array (ch4)
- CTM transcription is used as bootstrap for the FF transcription
- FF transcription is used as reference for RT'06

Transcription example



Distribution

- The CHIL data will be available to external partners through ELDA's catalogue
- ~40 hours of transcribed data
- Evaluation packages
  - Enable external players to run the evaluation offline
  - Data, tools, results of the official campaign

Summary

- Presentation of the lecture data for RT'06
- Multi sites data collection
- Same data used in many evaluations CLEAR/CHIL/RT
- Data publicly available through ELDA's catalogue